12-2014

# Measuring the Norm of Reciprocity on Data Sharing Practices: A Carrot or Stick Approach?

Crystal Pleake Sherline
*University of Tennessee - Knoxville*, csherli1@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Crystal Pleake Sherline entitled "Measuring the Norm of Reciprocity on Data Sharing Practices: A Carrot or Stick Approach?." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Communication and Information.

Suzie Allard, Major Professor

We have read this dissertation and recommend its acceptance:

Michael R. Kowtowski, Michael Olson, Carol Tenopir

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

# Measuring the Norm of Reciprocity on Data Sharing Practices: A Carrot or Stick Approach?

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Crystal Pleake Sherline
December 2014

## DEDICATION

I dedicated this work to my children, Liam Brier Sherline and Poppy Joy Genevieve Sherline. I never knew my strength until I became your mother. And to my nieces, Olivia Paige Pleake and Ciara Brooke Pleake. Thank you for allowing me to parent you, love you, and support you.

# ACKNOWLEDGEMENTS

I would first like to thank my committee members for agreeing to see me through this. Thank you, Suzie Allard, for encouraging me when I wanted to walk away. Thank you, Carol Tenopir, for seeing my potential early on. Thank you, Michael Olson, for stimulating conversation and interest. Thank you, Michael Kotowski. You have essentially held my hand through this process and I know how you feel about hand-holding. Seriously, I will forever be indebted to you and your family for your support, kindness, friendship and joy.

I would also like to thank the members of my cohort, who early on in the program helped encourage progress. Thank you, Rachel Jue Rui, Matthew Broaddus, Ilwoo Ju, Martijn Van Kelegom, and Rebecca Anderson. The cohort expands to the dungeon and all of you that has graced its walls.

Finally, I would also like to thank my parents and children. Thank you Roger and Sandra Pleake and Liam and Poppy Sherline.

# ABSTRACT

Based on the theory of the Norm of Reciprocity (NOR), this study is focused on an individual's data sharing behavior with respect to academic research by investigating their attitude towards data sharing and external funding. A measure was developed for data sharing, and the Adjusted Eisenberger Scale was attuned for measuring the Norm of Reciprocity. The measures were distributed by a random numbers generator to academic researchers at research intensive universities. The results show that NOR does not correlate with data sharing. There was also a negative correlation between scientists' willingness to share data and external funding. The results are inconsistent with the psychological theory.



**Keywords:** Data Sharing, Norm of Reciprocity, Social Norms, Information Science, Grant Funding, Academic Scientists, Research Intensive Universities, Externally Funded Research

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER I
# INTRODUCTION AND GENERAL INFORMATION

## Big Science, Bigger Data

In the early 1940s, the United States was at war. After the attack on Pearl Harbor, Hawaii by the Japanese, the U.S. could no longer remain neutral in World War II. Desperation required inspiration and research. Research on new weaponry, like the atomic bomb, was seen as a necessity in order for the U.S. to win the war. The necessity and perhaps desperation for scientific research and discovery, enabled collaborations from some of the best scientific minds of the time. An effort as large as winning a world war required collaboration and money. Within 6 years, weaponry was developed that would inevitably win the war (Hewlett & Anderson, 1962). Along with that weaponry, "Big Science" was born (de Solla Price, 1986). Science transitioned from smaller, institutional, research facilities, funded by individuals or corporations, to "Big Science" funded by government entities, such as the National Science Foundation (NSF) and National Institute of Health (NIH) (Galison & Hevly, 1992). "Big Science" required sophisticated experiments, apparatus, collaboration, and money (de Solla Price, 1986). The birth of "Big Science" coincides with technology and technological advances. Experimental apparatus eventually became computer driven and the data deluge began (Oettinger, 1965; Bello, 1960; Hearings, 1956). Data Deluge is defined as an overwhelming volume of information and the capacity to manage, use, and access that information (Hilbert 2011). According to Hilbert and López, "the world's technological information processing capacities are quickly growing at clearly

exponential rates" (2011). According to longitudinal study conducted at the University of California, Berkeley in 2000 and then again in 2003, there was approximately 160 Exabyte of stored information, with a projected compound annual growth rate of 57% per year (Lyman & Varian, 2003). Granted, not all of this information is scientific data; however a large amount are research data. How much scientific research data are out there somewhere and how does one access that data?

According to Dr. Francine Berman, chairwoman of the Research Data Alliance, no one can be certain about the amount or the availability of data, but according to Alan Blateckly, the director of advanced cyberinfastructure at the National Science Foundation, "data is the new currency for research" (Markoff, 2013). If data does indeed have value, in what way are data accessible to other researchers? How can one researcher obtain datasets from other researchers? "The casual approach for many scientists has been to 'stick it on my disk drive and make it available to anyone who wants to use it,'" explains Dr. Vinton Cerf, vice president of Google (Markoff, 2013). But what about scientists that do not make data "available to anyone who wants to use it'? With the sheer volume of data collected, why do some researchers and scientists remain hesitant about data sharing?

### *Platform Access and Institutional Repositories*

According to Corbyn, researchers decline to share data for many reasons, including, the time it takes to make the data available, lack of reward system for sharing

data, lack of consequences for not sharing data, or “the rewards systems aren’t there and neither is the stick,” and the lack of a “standardized format to make accessing data more efficient and feasible,” (Corbyn 2011). The problem of developing a standardized format for data archiving and retrieval still persists, in part because there is no new money for the new data deluge (Markoff, 2013). “Publicly accessible data requires a stable home and someone to pay the mortgage,” explains Dr. Berman (Markoff, 2013). Although, there are several institutional repositories (IR) that collect data from researchers and scientists, the IRs pose problems across platforms because there is a lack of a data sharing software standard. arXiv.org, CoRR, Repository 66, OpenDOAR, and Experimental OAI are a few IRs in operation and use. These IRs do not support multiplatform IRs and because of the lack of standards, sharing data across platforms, universities, and research institutes can be difficult, if not impossible (ROAR, 2013).

A first step in standardization of IRs is to start with an open access repository developed by a collaboration of librarians, scientists, potential public users, and university administration. Open access repositories are defined as non-exclusive access and retrieval of data on a searchable interface. In a study conducted at the University of Rochester an IR for archiving and accessing data sets was built in 2003. Six years after, the IR was unutilized (2009). Building the IR at the University of Rochester in conjunction with the scientists, computer scientists, and librarians still yielded no data sharing (Nelson, 2009). Although the University of Rochester’s IR was developed with the cooperation of scientists, researchers, and developers and training on the IR was available, the IR remains empty. Perhaps providing a standardized

platform for archiving and retrieving data sets is not the only variable in a researcher's decision to share or not share data?

### *Collaboration*

Within the last decade, there have been several debates concerning data sharing and collaboration. Within science policy circles, collaboration is seen as a progressive action and is often encouraged (Katz and Martin, 1997). Collaborations between universities and industry have also been on the rise (Katz and Hicks, 1997). Several studies have been conducted revealing the impact of collaboration, one specific study shows that collaborations with the biggest bibliometric impacts are authored by several people from multiple intuitions (Katz and Hicks, 1997) and from multidisciplinary fields (Kats and Hicks, 1995). "Growing collaboration is not only an expression of 'big science' but also a part of the globalization process in scientific research," (Glanzel and Schubert, 2004). However, no policy or procedure has been set into place to guide scientist on what collaboration involves. Some of the questions asked about collaboration are: what exactly is collaboration; who does what; how much credit should a collaborator get; if a colleague shares data, is that colleague a collaborator (Katz and Martin, 1997)?

Collaboration is a type of sharing (Katz and Martin, 1997). A number of studies have been conducted involving sharing behavior, including one study exploring the Theory of Reasoned Action, which is used as a prediction of behavioral intention by way of attitudes and subjective norms (Ajzen and Fishbein, 1980). In one study, the

researchers measured three factors of individual's attitudes toward knowledge sharing, including extrinsic motivation, channel richness, and absorptive capacity. They found that extrinsic motivation, which "refers to the performance of activities in order to attain some separable consequence" had no effect on sharing attitude. This led the researchers to conclude that "people basically do not care about what rewards they could attain by sharing their knowledge" (Kwok and Gao, 2006). They found an effect with channel richness, which indicated that the more transmission channels (information input), the greater the chance that an individual would share knowledge. The most profound construct in this study was that of absorptive capacity or learning process, which is defined as the "ability to not only acquire and assimilate but also to use knowledge" (Cohen and Levinthal, 1990). The researchers found that if there was a reciprocated knowledge exchange during the learning process, then both the recipient and the contributor are more likely to share information (Kwok and Gao, 2006).

### *The Norm of Reciprocity*

The Norm of Reciprocity, which states that people will return benefits for benefits, is therefore a better indicator of people's sharing behaviors. As a norm, reciprocity is used for persuasion, cooperation and for self-sustaining relationships (Gouldner, 1960). The theory of the Norm of Reciprocity states that the norm is universal but conditional, whereas, it is practiced across cultures, yet once the benefit has gone unpaid, the cycle stops; tit for tat, is self-perpetuated, as a debt repaid is repaid with interest, and so the cycle continues as the balance is incessantly reversed.

## Data Sharing and Reciprocity: The Connection

This section presents the process in which this research topic was formulated. Several iterations and questions arise when pondering the theory of the Norm of Reciprocity and its impact on scientific data sharing. It is inferred from the research that the Norm of Reciprocity could be examined with regard to data sharing. There are several assumptions with this assessment. First, the research assumes that Gouldner's theory is universal, meaning research scientists will prescribe to the norm and the normative theory. Secondly, it is also assumed that granting of a request, in this case, the receiving of data, will be seen as a benefit or favor. Thirdly, it is supposed that receiving public monies, i.e., a scientific grant, will also been viewed as a gift, or favor. Lastly, it is presumed that research scientist prescribe to the Baconion view of the scientific method, specifically, replication and independent verification of empirical research findings. It is with these assumptions and review of the literature that this study emerges.

### *Research Questions*

With theory, subject, and population in mind, the following research questions emerge:

*RQ1:* Are scientists who have a greater endorsement of the norm of reciprocity more likely to share data than scientists who have a lesser endorsement of the norm?

*RQ 2:* Are scientists that receive research support from a grant-funding source more likely to share data than those who do not receive support?

*RQ 3:* Does the number of publications of scientists affect sharing behavior?

*RQ 4:* Would incentives or penalties from funding sources endorse sharing?

### *Purpose of the Study*

The purpose of this quantitative, descriptive, cross-sectional survey is to examine scientists' behaviors and attitudes on data sharing and to test scientists' endorsement of the norm of reciprocity with regards to data sharing and to examine other variables that may affect scientific data sharing such as: 1. Time spent as a researcher. 2. Number of publications. 3. Type of funding.

***Table 1. Terms and Definitions***

| Term | Definition |
|---|---|
| *Hard Sciences* | Math, Astronomy, Physics, Psychology and Chemistry. Defined as a science that uses quantitative data. |
| *Data* | The values of qualitative or quantitative variables belonging to a set of items collected together for reference or analysis |
| *Data Sharing* | The practice of making data used for scholarly research available to other investigators |
| *The Norm of Reciprocity* | The norm of reciprocity states the expectation that people will respond favorably to each other by returning benefits for benefits |
| *Theory of the Norm of Reciprocity* | The theoretical foundation that the norm of reciprocity is a universal, conditional norm based on the understanding that people will respond favorably to each other by returning benefits for benefits and that people will not harm those from which they have received benefits. |
| *Renowned* | Being known by many people in a particular area of research |
| *Grants* | Endowments of research funding covering any funding for scientific researcher in which the researchers do not have to repay |
| *Grant Funding Institutes* | Government organizations that provide scientific researchers with endowments to conduct research |
| *Collaboration* | Collaborative associations between two or more research scientists and/or research institutions |
| *Requester* | A scientific researcher who asks for data from another scientific researcher |
| *Grantee* | A recipient of data |
| *Granter* | A contributor of data |
| *Carrot* | A reward |
| *Stick* | A penalty |

# CHAPTER II
# LITERATURE REVIEW

## The Carrot and the Stick: Rules and Expectations for Data Sharing Practices

Science has become big science. Experiments now involve intricate apparatus that cost billions of dollars. Private industry and government agencies fund these scientific endeavors. Funding agencies such as the National Science Foundation (NSF) and the National Institutes of Health (NIH) are the advocates of big science. With the backing of the U.S. government and tax payer dollars, these funding agencies have continued to promote science and scientific discovery. Along the way, there have been some rule changes. As technology advanced and science became more collaborative, funding agencies realized the need to implement data sharing practices (NSF, 2001; NIH 2003).

On April 1, 2001, The National Science Foundation (NSF) issued a press release announcing that the future of data sharing was going to change:

> NSF …expects investigators to share with other researchers, at no more than incremental cost and within a reasonable time, the data, samples, physical collections and other supporting materials created or gathered in the course of the work. It also encourages awardees to share software and inventions or otherwise act to make the innovations they embody widely useful and usable.

b. Adjustments and, where essential, exceptions may be allowed to safeguard the rights of individuals and subjects, the validity of results, or the integrity of collections or to accommodate legitimate interests of investigators (pg. 17).

The National Institute for Health (NIH) followed the lead of NSF and also implemented data sharing regulations for all of its grant recipients (NIH, 2003). Then, in February of 2013, a memo was sent from the Executive Office of the President , Office of Science and Technology Policy (OSTP) with the objective that "federally funded scientific research are made available to and useful for the public, industry, and the scientific community," (Holdren, 2013). With two of the largest grand funding institutes supporting data sharing practices, scientists began to understand that their data sets were now community property. Although the idea of data sharing seems productive in the advancement of science, some researchers remain skeptical about sharing their data sets (Ceci, 1988 and Tenopir et al., 2011). After synthesis and analysis of the data sharing research, several themes of why scientists do not want to share their data sets have emerged. Campbell et al found that researchers were concerned about several aspects of data sharing including: it requires too much effort; the grantor is protecting his/her rights or the research team's rights to the information for further publication (2002); and if the research is innovative and applied value, then there is monetary significance to the findings contained in the data sets and the initial researchers should have ownership rights (Bjaalie, 2007).

Douglass et al, in a similar study, found that government scientists do not share their data for several reasons, including insufficient time, lack of funding, lack of standards, lack of platform, and not having rights to make the data public (2014). Walters found with regard to faculty in academe, the reasons for not sharing data stem from not yet being published, no clear policy on acknowledgement, reciprocity, or collaboration, and data misuse (2014). Both studies look at populations that receive funding from institutes that require data management plans, yet sharing the data collected from these funded research projects are not publically available in most circumstances (Douglass et al, 2014; Walters, 2014). Walter's focus on NSF/NIH requirements and Douglass et al's focus on government scientists are all funded by public money and therefore data sharing is mandated, but to what extent? Data management may be a requirement to receive funding, but once the money has been distributed, what happens to the data? What happens if the researchers decided not to share their data? Although these agencies have demanded that data management become a part of every grant application, there is still a lack of policy for the mandate, especially with regard to rewards or penalties. There have been strides in the last decade to mandate data sharing, but in order to achieve the level of collaboration and transparency that should occur in scientific endeavors, stricter policies, practices, and penalties need to be implemented (Douglass et al, 2014).

The reason data sharing practices are changing in science is because of the common belief that science ought be transparent (Birnholtz &Bietz, 2003),

(Bjaalie, 2007), (Ceci, 1998), (Kaye et al, 2009), (Koslow, 2002), (Kraut et al, 1998), (Postle et al, 2002), Rashid et al, 2006), (Schofield et al, 2009), (Van Haouse et al, 1998), (Zimmerman, 2003, 2008). While most researchers believe that transparency is one of the key aspects of science and scientific discovery, scientists do not believe transparency is a problem in data sharing because results are available and published in the literature (Ceci, 1998).

By publishing results, other researchers are able to access the information on any findings in any subject domain. However, if data requests are being denied and data sets are not available, then the transparency of science is rightly being questioned. When scientists grant requests and share their data sets, one of the main reasons these scientists give for complying with such requests is for transparency of science (Ceci, 1998).

Another reason given for sharing primary data sets is for the promotion of scientific discovery (Birnholtz & Bietz, 2003), (Bjaalie, 2007), (Ceci, 1998), (Kaye et al, 2009), (Koslow, 2002), (Kraut et al, 1988), (Postle et al, 2002), (Rashid et al, 2006), (Schofield et al, 2009), (Van House et al, 1998), (Zimmerman, 2003, 2008). “Sharing data should make research more efficient and greatly facilitate our understanding,” (Koslow, 2002). When researchers in any domain, publish their findings in academic journals, they do not include their primary data. They present a particular phenomenon that they are interested in researching, a review of what is known about the phenomenon, and how they plan on looking at

the phenomenon, a methods section, analysis, and discussion. What are not present are other non-significant results or questions that were explored in the research. Upon publication, other researchers from the same domain read the journals and often times formulate their own ideas on how they would have approached the phenomenon. When these researchers find fatal flaws, they are more likely to try replicating the project to see what answers they receive. One of the first steps in replication would be asking for the primary data of the project initially carried out. One of the rules of scientific discovery is replication. It is through further study that discoveries are made. "Primary data will gain in value if it is put into the public domain once it has been analyzed and published. The combination of this new data with other data, and further analysis and correlation with other data, will lead to increased value, and new knowledge and understanding," (Koslow, 2002). The idea of increased knowledge value has led many institutions to create repositories for disseminating data and scholarly work.

Sharing data also allows other researchers to not only see how data was collected, but to also see what was explored and what was dismissed. By sharing primary data sets, other researchers in a domain can scrutinize the analysis of the data, note any incongruity, and reassess the data for maximum impact (Birnholtz & Bietz, 2003). By reevaluating data sets, new questions can arise and therefore new answers can develop.

Research today produces large amounts of data. After analysis and publication, the project is usually considered complete. When reviewing scholarly communication in journals, the data presented are never sufficient for reuse; therefore to ensure the ability of reexamination, data sharing has to become standard practice (Goodman et al, 2014). Their paper published in PLoS Computational Biology, give 10 rules to ensure good data sharing practices. Those rules are as follows:

1. Publish your data and encourage others to publish their data.
2. Share your data online but use a Permanent Identifier.
3. Conduct Science with Reuse in Mind
4. Publish Workflow as Context
5. Link your data to your publication
6. Publish your code
7. Establish how you want credit
8. Foster and use Data Repositories
9. Reward Colleagues who share their data properly
10. Be a Booster for Data Science. (Goodman et al, 2014).

In a 1999, fMRI Data Center was developed as an open-access data sharing platform in the neuroimaging community (Van Horn and Gazzaniga, 2002), where all data was available to "interpret, analyze, and replicate the deposited studies" (Mennes et al, 2013). In order to make the repository a success, the Editor-in-Chief of the *Journal of Cognitive Neuroscience*, required all contributors to the journal to deposit their data in the fMRIDC (Mennes et al, 2013). This mandate was followed by controversy. There were two major concerns. The first was that of technology/platform, as some of the data generated could be larger than 30 GB; how would storage and distribution work?

The second concern was that of losing a competitive edge in the research, especially with regards to promotion (tenure) and grant money (Mennes et al 2013). Vines et al also raise another concern and that is how can the researchers be assured that data will be available in a repository in the future (2013)> In a follow up study of the fMRIDC, it was determined that the initial mandate, which was controversial at the time, was exactly what made the fMRIDC a success.

> "Voluntary sharing of data may be the ideal, but many researchers find themselves very busy and, given the choice of how to spend their time, may not find the benefits to the community to be compelling enough to make the effort. Additionally, only having data from a "coalition of the willing" may not fully capture the breadth and depth of imaging experimental methods being applied across the field. On the other hand, sharing required as a condition of funding or journal publication can ensure a steady stream of data" (Van Horn & Gazzaniga, 2013).

There are several factors that determine whether or not an adoption of data sharing standards will succeed. Van Horn & Gazzaniga, in their longitudinal study, found that working closely with funding agencies and journal editors was a crucial element. Five other key aspects of data sharing include: good curation, multiple data sharing models, delivery of information in any form necessary, community engagement, and plan for the end of funding (2013). It was the implementation of the mandate 15 years ago that resulted in a successful model of data sharing. Since the implementation, the field of cognitive neuroimaging as seen advancements in several areas including: analytic methods, data mining, modeling, and visualization techniques (Van Horn & Ishai, 2007), thus promoting the purpose of science.

### *Repositories: Institutional and Subject*

> "A repository may be defined as a set of systems and services which facilitates the ingest, storage, management, retrieval, display, and reuse of digital objects. Repositories may be set up by institutions, subject communities, research funders, or other groups. They may prove access to a variety of digital objects, including peer-reviewed journal articles, book chapters, theses, datasets, learning objects, or rich medical files." (Pinfield, 2009, 165).

Institutional Repositories are generally institutionally specific. For example, IRs established in a university created so that faculty, staff, and students within an academic organization can share their work within their specific intellectual community instead of publishing to the "public webwide" open access repositories (Green and Gutmann, 2007). According to Pinfield et al, 2014, most repositories, on a global scale, are institutional repositories containing multidisciplinary, English-language documents and frequently exploit open-source software. The goal of institutional repositories is to collect as much research information as possible. In 2012, there were 2253 global repositories, where 1864 were institutional repositories, accounting for 83% of the total (Pinfield et al, 2014).There is also logarithmic evidence that the growth of institutional repositories is on the rise (Pinfield et al, 2014), with academic organizations prescribing to the adage, "if you build it, they will come." However, the success of an institutional repository is still dependent upon the faculty, staff, and students and their "willingness to contribute" to the repository (Green & Gutmann, 2007) (Lynch 2003). Foster and Gibbons completed a qualitative study on institutional repositories and found that these repositories are not

convincing enough for owners of the content to contribute. Authors and researchers want an institutional repository that would also benefit their needs, as well as provide a service to the interested public. The list of requirements compiled from the participants in the study include: easily accessible searches, preservation, linking capabilities, maintaining ownership of their work, no server maintenance, and no complications (Foster & Gibbons 2005). Therefore, just providing a platform for an institutional repository is not enough for information owners to supply their work. Because institutional repositories mature by means of voluntary deposits, capture rates will remain in short supply unless "institutions and funding agencies" create a mandate for sharing (Harnard, 2006).

A subject repository or domain-specific repository is the oldest form of information repositories (Xia, 2008). When information became digitized, a need arose to share the information with colleagues within the same research field (Green & Gutmann, 2006). Domain-specific repositories "hold collections of materials grouped by type, subject, or purpose and intrinsically support domain- or discipline-oriented research needs" (Green & Gutmann, 2006). The subject repository allows researchers to "reuse, repurpose, analyze, and recompile in teaching, learning, and research environments" (Green & Gutmann, 2006), which in turn, promotes the scientific method. In contrast to institutional repositories, where the goal is about quantity; the mission of subject repositories is to protect and preserve data in the long term. This preservation encourages more scientific

sharing, as the donators feel as though there work is being archived and protected (Bjork, 2013), and are therefore more likely self-archive their work.

Kling and McKim conducted an empirical study examining social knowledge sharing practices specific to discipline and found that there are differences in sharing between scientists of different domains (2000). The study also acknowledges that digital publication and preservation are (at the time) in an early period and in the future, there will be a shift from subject repositories to institutional repositories. According to a study conducted by Xia in 2008, physicists are the leaders of self-archiving in subject repositories. Xia goes on to infer, based on the theory of disciplinary culture, that physicist will be more likely to contribute to institutional repositories because of their past contributions to domain-specific repositories. However, findings of the research found a negative correlation between established self-archiving in subject repositories and the likelihood of self-archiving in intuitional repositories (Xia 200), meaning if physicists have already archived their work in a subject repository, they are less likely to re-deposit their work in an institutional repository.

Perhaps the time has changed, as there is a need and an effort for digital repositories, both institutional and subject specific to develop a partnerships with researchers in order to create a strategy that would benefit both contributors and users (Green & Gutmann, 2006). According to Pinfield:

> "The complex relationships between individual researchers, their organizations and subject communities, and their funders and governments, and the contributions all these actors make to repository development, need to be further explored in order to explain ongoing developments. Open-access approaches in general, and OA repositories

> in particular, have the potential to transform scholarly practice. The period since 2005 has seen considerable change in this area, with a repository infrastructure being established which is capable of playing an important role in scholarly communication. However, it is the next decade which is likely to reveal the extent to which these changes make a widespread and enduring impact on the scholarly community" (2014).

### *Data Sharing Practices and Behavioral Expectations*

Although there is a growing amount of literature concerning data sharing and data sharing practices, finding empirical research on the phenomenon is more difficult (Zimmerman, 2003). When researching data sharing and behavioral norms that may influence the practice of data sharing, the information available is more obscure. With regard to behavioral norms, in general and the norm of reciprocity, in particular, is especially important to examine in the context of data sharing because the norm is a powerful engine for motivating, creating, sustaining, and regulating the cooperative behavior required for self-sustaining social organizations (Axelrod, 2006) and is therefore the theoretical foundation of this study. Although reciprocity has been implicit in some studies involving data sharing (Tenopir et al, 2011, Zimmerman 2003, 2008), no measurement scales of reciprocity in regard to sharing data have been designed. With technological advances, increased construction of large, domain specific data repositories, scientific transparency, and requirement by grant funding institutions to share data, data sharing, or the lack thereof, is a phenomenon that is not going to simply go away. The theory of the norm of reciprocity and the power it carries

as a universal, behavioral norm (Gouldner, 1960), could help explain why some scientists participate in data sharing practices while others do not.

### *The Theory of the Norm of Reciprocity*

There are essentially two rules involved in the norm of reciprocity, first help those who have helped you and secondly, do not harm those who have helped you (Gouldner, 1960). Although the theory of the norm of reciprocity is thought to be universal, it is not considered unconditional. As Gouldner points out, “reciprocity connotes that each party has rights and duties,” indicating that the norm is only valid if both parties prescribe to the same helping behavior (1960, p.169). If for some reason a helping behavior is not reciprocated, then the natural self-perpetuating norm will cease to exist between the two parties. However, if helping behavior is reciprocated, it is never on an equivocal scale of the initial helped received and it is for that reason that the norm of reciprocity is self-perpetuating in “stable social systems” (Gouldner, 1960).

### *Compliance Benefits and Detriments*

Interestingly, for the scientists receiving a request for data, the costs of complying with this request may outweigh the benefits they potentially receive (Miller & Prentice, 1996). At the minimum, there are costs associated with the effort of locating and sending the data (or allowing the requester access if sending is not an option). Additionally, the sender will probably check the data’s

formatting and content to ensure no confidential information is shared, and make changes where necessary. Others who worked on the study may also be consulted before the data can be shared (Ceci, 1988). To the contrary, the benefits for the sender are limited: at best the receiver finds out that the sender drew the correct conclusions based on the available data (Campbell et al, 2002). Such findings tend not to be published, thus there is no potential for improved reputation. At worst, the receiver may identify flaws in the sender's published conclusions and expose those flaws, which may damage the sender's reputation and face (a potential cost) (Panagopoulos, 2010).

Sharing data thus appears a seemingly irrational act, of which personal costs outweigh potential benefits. Research has shown other examples of humans' acts that come with costs but few tangible benefits. For example, Panagopoulos found that voting is an act for which where social norms can override the disparity between individual costs and benefits (2010). Aronson, an evolutionary psychologist, believes that there are survival mechanisms for helping or altruistic behaviors (2007). By helping others, we are assuming that in the future others will help us and by continuing the self-perpetuating norm of reciprocity, we will continue to help and be helped. These helping behaviors enable the survival mechanisms to stimulate a survival mechanism (Aronson, 2007). For this mechanism to work in a data sharing context, the scientist must be aware of the social norm that proscribes data sharing. Three primary sources of data are evaluated to construct social norms: observable behaviors, direct and

indirect communication, and knowledge of the self (Miller & Prentice, 1996). The source of the norm impacts how it will operate, i.e., use of appeals to gain compliance, which could be a direct or indirect communicative effort.

### *Injunctive and Descriptive Norms*

Cialdini, Reno, and Kallgren distinguish between two types of social norms: injunctive norms and descriptive norms (1990). Injunctive norms are based on what is morally approved or disapproved conduct. They thus refer to what ought to be done. Descriptive norms are based on what is typical or normal behavior. They thus refer to what most people do, and that behavior is imitated. Although what is morally approved is usually what most people do, this is not always the case; injunctive and descriptive norms at times may misalign.

When data sharing is an injunctive norm for scientists it specifies what a good scientist ought to do. When data sharing is an injunctive norm for scientists it specifies what other scientists do as a guide for behavior. As this type of norm is learned through observable behaviors, the scientists for whom data sharing is an injunctive norm have been exposed to acts of data sharing, either by seeing their colleagues share data or by benefiting from receiving data upon their own request. Cialdini, Reno, and Kallgren found that each of the two types will only impact behavior when it is salient at the moment (1990). As it is not clear which type of norm is more salient in the context of data sharing, this study will appeal

to each type to test the overall effect of appealing to social norms and to test against each other the effects of appealing to the two types.

Scientists do their work (conducting research, presenting findings and conclusions) within the context of the scientific community. In addition to the rules that society at large sets for behavior, the scientific community has evolved its own set of guidelines for appropriate behavior that indicates how scientists are expected to conduct their work (Ceci, 1988). The scientific community is guided by one objective: the quest for knowledge (Zins, 2007). To move toward this objective and achieve valid knowledge, procedures, measures, and conclusions of studies are then submitted to field-specific journals, where they are subjected to the peer-review process. The findings are published and then read by other experts in the field. When these scientists read scholarly journals, they become aware of what other scientists are working on within their domain. When published findings question or negate previous findings, scientists become interested in re-examining the experiment, data, and/or analysis. Science should be transparent. Part of the scientific process is being able to recreate a study, comparing results with similar studies, and gathering data for analysis and conclusions. This arduous process is part of the scientific method. Conducting experiments is time consuming and costly. In some cases, the entire study may not need to be replicated; perhaps the analysis needs to be reviewed. In these studies access to the original data could provide immediate insight to questionable or perhaps revolutionary findings. Within the scientific community,

the guideline is that such access ought to be granted. Thus, scientists are expected to share their data when data is requested. Such a guideline for appropriate behavior set in a community is conceptualized as a social norm (Cialdini, 2007).

### *To Share or Not To Share*

Traditionally, collaborations happened in single laboratories, where scientists and scientists-in-training would share instrumentation and ideas, however with the introduction and use of electronic documentation, laboratory collaboration has shifted from confided spaces to less physical addresses (Birnholtz & Bietz, 2003). Although the ability to disseminate and house large sets of primary data are now a reality, scientists have often been hesitant to share their data (Campbell et al, 2002). According to Birnholtz and Bietz (2003), there are three over arching reasons why data sharing is not a universal practice:

1. scientists are not willing to share
2. there are problems locating data sets
3. there are problems when determining how one uses shared data

As discussed by Birnhotlz and Bietz (2003), Campbell et al (2002), Ceci (1988), Chui (2006), and Zimmerman (2003), sometimes scientists are just not willing to share their primary data sets. There are many reasons scientists give

for not wanting to disseminate their data, but the fundamental reason that has reoccurred in the literature is simply that the scientists do not want to. There is often a feeling of absolute ownership between the scientists and the data sets (Zimmerman, 2008), but there are other explanations of why scientists are not complying with data requests.

Hauesser, in her comparison of academic and industry data sharing, found that data sharing is dependent upon two variables: the culture of the institution and whether or not the information requested has economic value (2009). Another issue that has been previously stated about why scientists do not grant requests of their data sets involves the actual storage and location of the datasets (Birnholtz & Bietz, 2003). Although there have been considerable advances in technology, including the ability to store large data sets, there remains problems with access and dissemination of information (Stolte et al, 2003). As requirements for data management plans emerge, the ease of accessing informational data sets will be established and in turn, scientist will be discouraged to assert that their data sets are somewhere but they cannot remember where or how to access them. Perhaps with the development and acceptance of repositories, both institutional and subject specific, the need to remember and or ease of access can be alleviated?

The final identified concern about not sharing data sets involves use. Scientists are hesitant to share their primary data because the requestor of the data does not have to disclose the intended use of the data (Postle et al., 2002)

(Schofield et al., 2009) (Van House et al, 1998). “Once in the possession of a data set, understanding it requires knowledge of the context of its creation,” which means if a researcher does not understand all the procedural knowledge, the data sets could be of no use (Birnholtz & Bietz, 2003). Campbell et al also address the issue of use and determine that when scientists runs a particular experiment, they have a particular question in mind; experimental design, data collection, and analysis all depend on the question being asked (2002). When another scientist wants to run a meta-analysis on a primary data set, the questions being asked are typically not the same questions the data set were initially collected to answer, i.e., predicating post-hoc hypotheses.

The questions still remains, why do some scientists regularly share their data sets while others never share their data? How does the norm of reciprocity influence the scientists’ sharing behavior? Is requesting primary data sets like bargaining for a first born child?

Freedman and Frasier found that compliance was more likely to be granted from a requestee if a small favor was initiated first (1966). The foot-in-the-door technique has greater compliance rates than asking for large favors first. The idea is that a relationship is developed after an initial small favor is requested. If the small favor is agreed upon, then the relationship is open to larger favors (Freedman & Frasier, 1966). However, Cialdini et al found that the inverse is also true (1975). If a person first asks a very large favor, perhaps for participation in a large-scale research project, but the large favor is followed by a

less demanding favor, like access to a primary data set, then the requestee is more likely to comply with the smaller favor (Cialdini et al, 1975). Cialdini et al also measured controlled compliance by asking only for a small favor, however compliance was fifty-percent higher when a large favor was followed by a small favor (1975). If scientists believe that they have ownership rights to their primary data sets, then asking for their data on an initial contact may result in less compliance. As the data shows, it would be wiser to either ask for a small initial favor, i.e., contact information on someone who does similar research, and then follow that first request with the larger request of data sets or ask for a larger favor, i.e., participation in a large-scale, time consuming study, and immediately ask for a smaller favor, the primary data set.

***Status, Power, Face***

Marwell & Schmitt found that there are sixteen types of compliance gaining techniques and within those techniques, some variables that affect compliance (1967). One of the major moderators on compliance is face or power of the requestor, i.e., status. Face or power is described as how much influence one has on a particular domain (Iamnitchi et al, 2002; Tenopir et al, 2011). The theory of social power states that power is developed over time and that social influence is dependent upon communicative interactions (French, 1956). In order for a scientist to be considered renowned, there must be consensus by other scientists in that area of discipline (Lippitt et al, 1952). If a renowned scientist

requests a data set from a colleague, he is more likely to receive his request simply based on his reputation as a renowned scientist (Deutsch, 1955). A requestor with high power, having a respectable scientific reputation, is thought to be more thoughtful, critical, engaging, serious, forthcoming, scientific, and connected (Bourne & Barbour, 2011). According to Lippitt et al, "members with high attributed power receive more deference behavior from other members and initiate more social influence than do low power members" (1952, pg. 59). Conversely, Regan found that a person of high social power, perhaps renowned in his domain, is less likely to comply with data requests (1971) (Mui et al, 2002), indicating that those who are more renown in their research field may be less inclined to share data.

### *Data Sharing and the Norm of Reciprocity Connection*

Now that the major grant funding institutes have regulated data sharing practices and data management plans, scientists will be obliged to share and disseminate their primary data sets. What is most interesting about the norm of reciprocity is that the behavior is dictated by society. In science society, data sharing practices are becoming common procedure and within that system, the norm of reciprocity will also become customary. Although some scientists are opposed to sharing, with the grant funding agencies requiring allocation of all data, how will the norm of reciprocity be adapted to by the scientific community?

With science opening up to new levels of transparency only made possible by technological innovation, science could be on the brink of a revolution?

### *Research Questions for Analysis*

$RQ_1$: Are scientists who have a greater endorsement of the norm of reciprocity more likely to share data than scientists who have a lesser endorsement of the norm?

$RQ_2$: Are scientists that receive research support from a grant-funding source, more likely to share data than those who do not receive support?

$RQ_3$: Does the number of publications of scientists affect sharing behavior?

$RQ_4$: Data are more likely to be shared if the requester is renowned in the field than if the requestor is not renowned.

# CHAPTER III
# METHOD

## Measures and Method

This section describes the methods used to conduct the study, and presents and explains the research questions. The methods discussion includes descriptions of the population randomly sampled for participation in the study, the operationalization of the research questions, the way in which the study was disseminated, and the techniques employed for data analysis.

### *Definition of the Population*

In order to test the theory of the norm of reciprocity in the context of data sharing, data were collected from research scientists conducting research in the hard sciences at 25 Research Intensive Universities across the United States. Theses Research Intensive Universities (RIUs) were listed in the Carnegie Classification of Institutions of Higher Education (2014). The foundation classifies 108 universities as RUIs with very high research activity. The potential RIUs were then narrowed by institution type and private universities, which numbered 35, were discarded. The remaining 73 public RUIs were added to a random numbers generator (Babbie, 1990) where each RUI had the same chance of being selected as the others. After determining which 25 RUIs were selected, the researcher consulted the faculty directory of several departments at the 25 RUIs.

The departments were in the hard sciences: physics, chemistry, microbiology, astronomy, and in psychology. The researcher created a spread sheet with the email addresses of all researchers at the 25 RUIs conducting research in the scholarly areas listed above. Based on a salary study conducted in 2012, there are approximately 1.5 million faculty and teaching members at American universities, of that population, approximately 500,000 are faculty members of RUIs. The population can then be narrowed by discipline and of the 500,000 faculty members at RUIs, 100,000 are faculty researchers in science (Lederman, 2012). Thus making the sampling factor $\frac{n}{N}$, whereas *n*=1400 and *N*=1400, $\frac{1400}{100000}$ = 1.4 % of the total population was contacted to participate in the study.

The cross-sectional survey was distributed to 1400 possible participants by random numbers table (Babbie, 1990) on 21 January 2014, after receiving IRB approval. Potential participants were identified via universities' faculty directory and an explanatory email (see Appendix D) was sent to the individual research scientists and faculty members, along with a unique survey link (see Appendices A, B, and C). This plan resulted in a sample consisting of n=84 faculty members from 25 RUIs. Recipients were neither compensated nor penalized for participating.

### *Operationalization of the Research Questions*

This section enumerates the concepts surrounding the research questions, and the variables that convey them, connecting the concepts to the

indicator items from the measurement instrument. There are two major concepts of the study. The first is endorsement of the norm of reciprocity. The second is scientific data sharing. There are also three minor themes of the research. The first is time spent as a researcher. The second is number of publications over the researchers' careers. The third is whether or not the researchers' work is grant funded. All of these concepts areas are measured directly by one or more questions on the survey instrument.

Research question 1 states: Are scientists who have a greater endorsement of the norm of reciprocity more likely to share data than scientists who have a lesser endorsement of the norm? This question is principally a summary of the overarching theme of this research. The norm of reciprocity is a social norm, or a set of rules that a group or society deems appropriate or inappropriate in terms of values, beliefs, attitudes, and behaviors (Marshall & Johns, 2009). In order to measure the participants' endorsement of the norm of reciprocity, Eisenberger's Adjusted Scale was employed. Eisenberger's Adjusted Scale is a 20 item, 7 point, Likert-type scale, with potential responses ranging from strongly disagree (1) to strongly disagree (7) was used to indicated how much respondents agree or disagree with the ideas expressed by each item statement. All of the statement used to measure the independent variable express attitudes associated with the norm of reciprocity.

Research question 2 asks: Are scientists that receive support from grant-funding sources, more likely to share data than those who do not receive support? The research question was addressed by asking an open-ended single item question asking if the participant was receiving grant funding. That answer was correlated with the scores on the index calculated from the indicators on Eisenberger's Adjusted Scale, which measured the participants' endorsement of the norm of reciprocity.

Research question 3 asks: Does the number of publications of research scientists affect sharing behavior? This research question is closely tied to research question 2, where experience was measured by years in the field; however this dependent variable is measured by 1 open-ended question asking participants to estimate how many publications they have produced since entering their research area. This measure was then calculated for correlation with the participants' scores on the index of Eisenberger's Adjusted Scale.

Research question 4 states: Is data more likely to be shared by a renowned researcher? This question was posed in the Data Sharing Scale, item number 9, which asks, "Pat should share data with researchers who are renowned in Pat's area of research." According to previous research conducted by Tenopir et al, 2011, scientists that are more renowned or experienced are more likely to share data than those researchers who are less renowned. It was inferred in that particular study that being renowned in a research area comes

toward the end of a career, when competition for publications, promotions, etc, are less advantageous (Tenopir et al., 2011).

This question addresses the potential variable of “experience” on data sharing. In order to determine how experienced a researcher was, one two open-ended questions were asked. The first question was number of years researching in your field. The other question posed was number of publications. It is understood that the number of years in a particular research area and the number of publications would have a positive relationship meaning, the longer a researcher has been conducting research, the more publications that researcher would produce. Therefore it is understood that with publications over time, there would be growth. The numbers of publications are also an indicator of how renowned a researcher is in a particular research area. After data collection, it was determined that the participants’ responses to number of years in the field were an accurate measurement of experience and therefore no analysis of variance test executed. The response from number of years in the field was correlated with scores on the index of Eisenberger’s Adjusted Scale to determine the relationship.

## Procedure

Participation was solicited by an invitational email where participants were asked to complete a questionnaire (See Appendix D). The invitation email also contained a unique link to the survey, using Survey Monkey©. One week later, a

reminder was sent to those who had not yet participated in the survey. This was repeated for four weeks, after which the study was closed.

### *Instrumentation*

In addition to a measure containing several demographic items (Appendix C), such as age, race, research area, and years as a faculty member, number of publications, the questionnaire was composed of two separate measures: the Adjusted Eisenberger Scale (AES) Data Sharing Scale (DSS). The AES was administered to rate the participants endorsement of the norm of reciprocity and the DSS which was developed to measure data sharing attitudes, was created by the researcher for this study. The original Eisenberger scale which was developed in 2004 consists of 24 items with response scales ranging from *strongly disagree* (1) to *strongly agree* (7). In past research, the measure has been found to be reliable with reliabilities ranging from α=.81 to α=.91(Eisenberger et al, 2004; Boster et al, 2005). Van Horn et al, showed that the global assessment of reciprocity is valid in terms of convergent and construct validity (2001), but suggest using self-rated scales that measure reciprocity with a single item because the scales "are conceptually closest to the phenomenon to be explained and they measure the balance of exchange" (Väänänen et al, 2005, pg. 179). The AEG consists of 20 Likert-type items with the scale continuum ranging from *strongly disagree* (1) *to strongly agree* (7) (see Appendix A). The second measure, the DSS, consists of 24 Likert-type items developed to

measure attitudes and beliefs about data sharing practices (see Appendix B). Figure 2, shows how the developed measures correlate with the hypotheses and research questions. The data collection involved creating a Web-based survey and administering it online using Survey Monkey© (Nesbary, 2000; Sue and Ritter, 2007).

## Table 2: Variables, Research Questions & Items on the Survey

| Variable Name | Research Question | Items on the Survey |
|---|---|---|
| Independent Variable 1: Norm of Reciprocity | Descriptive Research Question 1: Are scientists who have a greater endorsement of the Norm of Reciprocity more likely to share than scientists who have a lesser endorsement of the Norm of Reciprocity? | See Questions 1-20; Eisenberger's Adjusted Scale |
| Dependent Variable 1: Data Sharing | Descriptive Research Question 2: Why do some scientists share data while others do not? | See questions 1-24, Data Sharing Scale. |
| Control Variable 1: Area of Research | Descriptive Research Question 3: Does the area of research affect data sharing? | See Question 2, Demographics, Appendix C |
| Control Variable 2: Experience as a researcher (time) | Descriptive Research Question 4 Does experience of the researcher, defined as the amount of time and research a scientists has conducted, affect data sharing? | See Question 3, Demographics, Appendix C |
| Control Variable 3: Experience as a researcher (publications) | Descriptive Research Question 5: Do the number of publications of scientists affect sharing behavior? | See Question 5, Demographics, Appendix C; |
| Control Variable 4: Funding Source/Carrot or Stick? | Descriptive Research Question 6: Do funding agencies proved incentives and /or penalties for data management plans/data sharing practices? | See Question 4, Demographics, Appendix C; See Questions 1, 2, 4, 20-24 on Data Sharing Scale. |

## Table 2: Continued

| Variable Name | Research Question | Items on the Survey |
|---|---|---|
| Control Variable 5: Are you part of a minority group or under-representative group? | Descriptive Research Question 7: If the norm of reciprocity is universal, what will sharing practices look like in underrepresented groups? | See Question 6, Demographics, Appendix C. |

# CHAPTER IV
# RESULTS

In this section the results of the analysis will be displayed and discussed.

## Results

The sample consisted of n=84, or a 6% response rate. A response rate of 6%, although not ideal, is considered to be valid because the sample was randomly drawn. The sample is large enough to allow for decent, preliminary analysis. The sample consisted of research faculty, participating in quantitative research at RUIs in the academic fields of physics, biology, chemistry, and psychology.

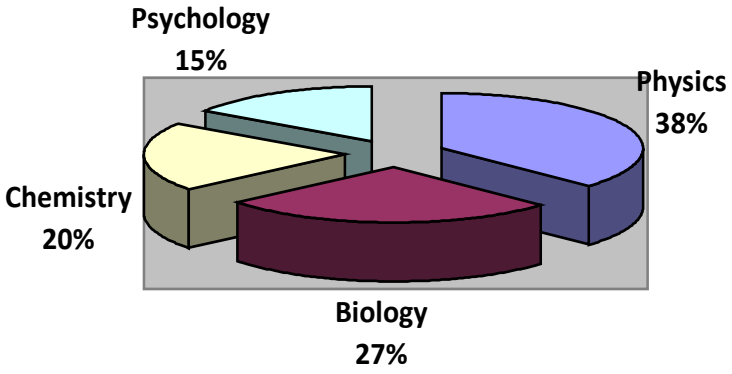


**Figure 2. Participants by Research Area**

Research faculty at RUIs are continuously conducting research and some of that research is time sensitive, therefore social science research participation

is not viewed as being crucial. The initial survey was distributed on 21 January 2014, which was also at the beginning of the spring semester for most faculty members, so an optional survey was not a focal point. There was no reward or punishment for participating, and with no incentive to contribute, most chose not to do so.

## Descriptive Statistics

The participants consisted of 29 % female and 71% male, which is representative of the sample of faculty members in hard sciences at RUIs. Participants ages ranged from 29 to 81 years old ($M$=50.68, $SD$=12.24). The participants reported number of years working in academe as $M$=18.92, $SD$=12.99, with a range from 6 months to 47 years. Participants also reported number of publications during their tenure as $M$=80.18, $SD$=91.22, with a range from 0-400.

## Measure Reliabilities

After determining reliabilities for each of the two scales, it was apparent that participants experienced end of survey fatigue, as the measures at the end of the instruments were causing the reliabilities of the measures to plummet. Three items were dropped from the NOR measure resulting in a 9 item scale with

a reliability of α=.81. Twelve items were dropped from the DSS measure resulting in an 8 item scale with a reliability of α=.92.

Table 3: Norm of Reciprocity Reliability Scale

| Mean | SD | Min-Max | Skew | Kurtosis | Cronbach's α |
|---|---|---|---|---|---|
| 3.90 | 1.04 | 1.29-6.14 | -.07 | .03 | .81 |

Table 4: Data Sharing and Demographics Index Scale

| | Mean | SD | Min-Max | Skew | Kurtosis | Cronbach's α |
|---|---|---|---|---|---|---|
| Data Sharing | 5.34 | 1.0 | 2.92 -7.0 | -.15 | -.38 | .92 |
| Age | 50.68 | 12.24 | 29-81 | .22 | -.57 | |
| Time in Acad | 18.92 | 12.99 | .5 – 47 | .46 | -.79 | |
| # of Pubs | 80.81 | 91.29 | 0-400 | 1.74 | 2.82 | |

For research question 1, which states: scientists who have a greater endorsement of the norm of reciprocity are more likely to share data than scientists who have a lesser endorsement of the norm, a NOR and DSS index

was compiled and a simple correlation was calculated, $r$=0.02, ($p$=*NS*). There was no significant correlation found between degree of endorsement of reciprocity and data sharing.

Table 5: Descriptive Statistics and Intercorrelations for Study Variables

| | | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|---|
| Data Sharing | (1) | -- | | | | | |
| NOR | (2) | .07 | -- | | | | |
| # of Pubs | (3) | .11 | .29* | -- | | | |
| Grant Funded | (4) | -.24 | .02 | .20* | -- | | |
| Time in Acad. | (5) | .09 | -.33* | .62* | .06 | -- | |
| Age | (6) | .07 | -.29* | .59* | .02 | .92* | |

Research question 2 states: scientists that receive research support through a grant-funding source are more likely to share data than scientists that do not receive support? Bivariate correlation was calculated for the NOR index

and grant funding. No relationship was found between data sharing and/or endorsement of the norm of reciprocity, number of publications, age, or time as an academic researcher. Multiple regression analysis was used to test if grant funding predicted participants' sharing behavior. The results of the regression indicated the two variables had a negative impact upon each other, $R=0.32$, $F(5, 68) = 1.59$, NS, where grant funding had a negative correlation with data sharing, $R^2 = -0.29$, $p<0.05$. This indicates that scientists who receive grant funding are less likely to share data than scientist that do not receive grant-funded support.

Research question 3 states: Do the number of publications affect sharing behavior. Because of the results of research question 2, there is no significant relationship between number of publications and data sharing, therefore no other analysis was conducted.

Research question 4 states: Data are more likely to be shared if the requester is renowned in the field than if the requestor is not renowned. Being renowned is determined by number of publications and the number of years working in the academic area. No significant correlation was found between data sharing and being renowned as a researcher.

### *Post-Hoc Analysis*

After examining the results, it was recommended to explore more analysis in regard to the norm of reciprocity scale with demographic information provided by the participants.

One particular item of interest was whether or not discipline affected the endorsement of the norm of reciprocity. In order to analyze the data, reported discipline was re-coded into four areas, disregarding specialized fields of physics, biology, chemistry, and psychology. For instance, if a participant reported the area of discipline as 'particle physics,' that response was re-coded as 'physics'.

Regression analysis was performed for each dependent variable to predict the overall endorsement of the norm of reciprocity. The full model is not statistically significant [F= 5.43, (5, 52) = .57, p=0.69]. There was not a significant effect of discipline on the endorsement of the norm of reciprocity at the $p<.05$ level for the four groups, [F (3, 77) = .49, p=0.69]. Discipline area is not a predictor of the norm of reciprocity. Physicists (M=3.72, SD=1.01), Chemists (M=4.05, SD=1.29), Biologists (M=4.00, SD=0.60) or Psychologists (M= 3.85, SD=0.95) in this sample, have no effect on the endorsement of the norm of reciprocity.

Although the full model was not significant, previously it was found that the correlation between number of years since receiving a doctoral degree and endorsement of the norm of reciprocity had a relative and positive, meaning that individuals who were more recent Ph.D.s had a higher endorsement of the norm of reciprocity. Experience was reported by two different variables in this study. The first variable was time since receiving a doctorate degree; the second variable to measure experience was number of publications. In order to explore whether experience effects one's endorsement of the norm of reciprocity, a one-

way between subjects ANOVA was conducted. There was a significant effect of time since receiving a doctorate degree on the endorsement of the norm of reciprocity at the $p<.05$ level, [$F(4, 76) = 3.72$, $p=.008$]. Post hoc comparisons using Tukey HSD test indicated that mean score for the researchers having received a doctorate within the last 5 years ($M=4.23$, $SD=1.22$) was significantly different than the researchers who received a doctorate degree 25 years ago or longer, ($M=3.39$, $SD=0.68$). These results suggest that in this sample, time since receiving a doctorate degree has an effect on the endorsement of the norm of reciprocity. Specifically, the results suggest that the newer researchers have a higher endorsement of the norm of reciprocity than older researchers.

Another way experience was measured in this study was by the number of publications the participants reported. A one-way between subjects ANOVA was conducted to compare the effect of number of publications on the endorsement of the norm of reciprocity. There was no significant effect of the number of publications on the endorsement of the norm of reciprocity at the $p<.05$ level, [$F(5, 67) = 1.43$, $p = 0.23$). Taken together, these results suggest that number of publications has no effect on the endorsement of the norm of reciprocity.

These post hoc analyses help to determine what variables in the sample could affect endorsement of the norm of reciprocity and after conducting the analyses, it appears that number of years since receiving a doctorate degree is the only variable of significance.

Because experience of researcher, measured by time since receiving a doctorate degree, had a significant effect on the endorsement of the norm of reciprocity, it was determined that post hoc analysis on data sharing amongst the same variables: number of publications, age of researcher, and number of years since receiving a doctorate degree should be examined.

Taken together, these results suggest that in this sample there is no relationship between discipline and number of publications or time since doctorate degree and data sharing practices, as the full regression model fit was not significant.

# CHAPTER V
# DISCUSSION AND LIMITATIONS

This study is a preliminary report on testing the theory of the Norm of Reciprocity on scientific data sharing. The data in this study have shown that the theory of NOR has no correlation with data sharing practices. It also suggests that endorsement of the norm of reciprocity is not dependent upon area of research (discipline), number of publications, nor time since receiving a doctorate degree.

After conducting this study, a myriad of questions emerge. Gouldner's theory of the norm of reciprocity claims that the norm is universal and imperative for survival (1960). If that is the case, then why is data that is funded by external organizations that require data management plans not available to the public? Why are scientists that receive external monies and put in place a data management plan, as required in the application procedure, not sharing? If reciprocity is not a variable that needs to be considered in data sharing, then what are the other variables impeding data sharing?

Goudner's theory of the norm of reciprocity claims that the social norm is universal. He postulates that throughout history, humans have relied on reciprocity to develop alliances and strong group ties (1960). As seen in this study, researchers in academe do not have a strong endorsement of the norm. After reading copious amounts of literature in social and evolutionary psychology, as well as in communication studies, an idea about not sharing arises. If

reciprocity is a favor for a favor, perhaps data sharing is not seen as a favor. Perhaps because grant funding is not seen as a favor, the scientists receive grants may not feel compelled to share their data. Another idea of limitation arises when thinking about first-hand gifts. Although Gouldner does not emphatically declare that his theory only works in first-hand instances of tit-for-tat, it is assumed that person-to-person gifts or favors are implied. If that is the case and the theory of the norm of reciprocity is limited to first-hand exchanges. This variable of direct contact is one possible explanation for the lack of sharing behavior seen with the sample population.

Although this study has shown no relationship between the endorsement of the norm of reciprocity and data sharing, what this study has shown is that there is a negative correlation between grant funded research and data sharing. What does that mean? Simply stated, if a researcher is receiving external funding, he is less likely to share his data. If researchers who are not externally funded are more likely to share data than those who are, how can external funding agencies encourage and, in fact, demand that researchers give public access to that data? Perhaps external funding is not seen as a favor, but is an expectation, with that in mind, why would recipients share their data?

What is interesting about this finding is its potential implications in scientific pursuits. With apparatus, materials and space becoming more expansive and expensive, and scientific cooperation obligatory, the lack of willingness for scientists to share data will have negative impacts on all scientific

efforts. The research lifecycle, as presented by the Joint Information Systems Committee (JISC), states that data sharing is a part of the research process (2010).

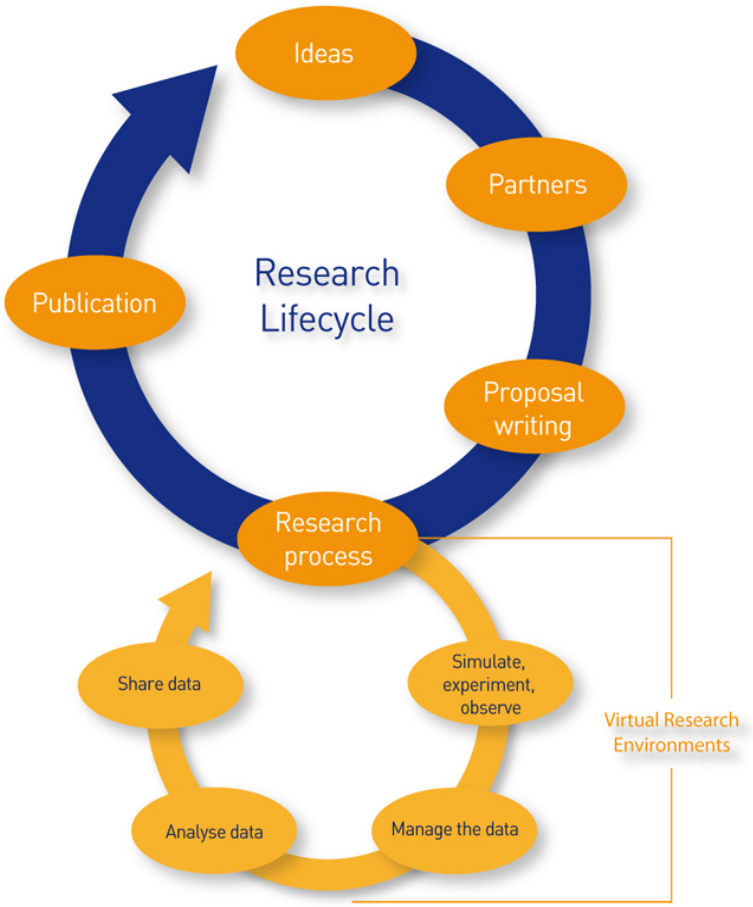


**Figure 4: Research Lifecycle**

JISC has one central principle and that is that "publicly funded research data should be openly available to the maximum extent possible" (2010). This current research study suggests that exactly the opposite is happening. Why?

A study by Arzberger et al suggests that there are several reasons data is withheld, accessibility to a hosting platform is the number one reason (2004). A publically accessible platform for all research and datasets seems in the distant future, however in the meantime, open access journals, like PLoS has implemented specific data sharing guidelines (Tenopir et al, 2011). In a 2009 study conducted by Savage and Vickers, 10 PLoS published authors were asked share their datasets. Out of the ten researchers contacted, only one researcher sent an original dataset.

Another reason cited for not sharing data are concerns about future research and publication using an existing dataset (Savage and Vickers 2009). The level of competition for public monies is extraordinary and the ability to receive a grant based on past and present research is necessary. Data withholding could be due to the competitive nature of funded research, in which case, the way grants are applied for, allocated, and then distributed, must be reviewed. Birnholtz & Beitz found that data goes unshared because researchers are proprietary over their data (2003). Ideas are generated by human beings, and although the government funds a majority of scientific research, the researchers are producing ideas. Data is the manifestation of those ideas, after design and testing has occurred. The government may have the right to request the data sets from their funded researchers, but what happens to the innovation, the idea? When data is shared, does the research idea as well as the implementation of

the idea, the method, the apparatus, the results, as well as the analysis, become public domain information?

Perhaps one way to ensure data sharing for researchers using external monies is to provide a platform to dump the data. If a platform is developed for externally funded research, which is accessible to all people, there would still have to be consequences for withholding data or researchers would not take the time and effort to upload their datasets. Seemingly, data sharing with regard to government funding could benefit from implementation of penalties for withholding, the stick approach.

While conducting this study, a memorandum was disseminated from the Office of Science and Technology Policy (OSTP). This memorandum will change the data sharing practices of every researcher receiving government monies. The memorandum alters data sharing policies initiated by NSF in 2001. It states:

> "The Administration is committed to ensuring that, to the greatest extent and with the fewest constraints possible and consistent with law and the objectives set out below, the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data."

With this new policy in place, all federal funding agencies have to have guidelines and procedures for open access data repositories and dissemination.

## Limitations and Further Research

This study is small stepping stone to discovering and refining data sharing practices. There are many limitations to this research. Firstly, the sample size is very small. It can be justified by the represented sample, of academia faculty. Some populations are harder to reach than others and for that reason a random sample was employed. Although the resulting participants represent a very small section of the entire population of interest, the study is valid. Because of the small sample size, it was irrelevant to test data sharing practices amongst groups of researchers in differing fields. The hypothesis that, for example, physicists share more data than chemists, would add general knowledge to the social science world. When the discovery of why some researchers share their data while others do not emerges, stakeholders can use that information to persuade the non-sharers to share.

Another limitation to the study is the actual measurement instruments. The Eisenberger's Adjusted Scale was adjusted to read in the third person. The initial thought of the researcher was participants will be more truthful if they do not take the measures personally. In hindsight, the researcher is unsure if that was beneficial. A pilot study of that scale would have beneficial however given the special population, it would have been difficult to test on a representative sample. The Data Sharing scale, too, could have benefitted from a pilot study to test the

measures and perhaps could have been disseminated to a group of available participants, i.e., graduate students, to test validity and reliable of the scale. A pilot study using a small population of graduate students, presumably, future academic researchers, would have given some insight into the depth and breadth of data sharing. It could also raise new questions about data sharing, such as: Are there generational differences in behavior with regard to data sharing? Are there generational differences in endorsement of the norm of reciprocity? Are less senior researchers more comfortable with the idea of open access repositories for data?

This study is not comprehensive and much more research needs to be done. Research on data sharing should be conducted and policies for public monies should be reviewed and revised. It is only then the promotion of science and scientific discovery can really be public. With new policy being enacted, as of 2013, the act of data sharing by researchers funded by government monies should no longer be part of the discourse. If these institutes provide clear guidelines on where, how, when, and what to do with research data and uphold penalties and rewards, then all federally funded research will become open access. Open access data will enable further innovation, curation, preservation, and analysis. A longitudinal study could determine if the policy change in data sharing is effective.

## Journals and Data Policy

Beginning in 2003, the National Institutes of Health (NIH) enacted data sharing plans for all large, grant funded research (NIH, 2003). Although a data sharing policy was put forth, there have been problems with implementation on the part of the researchers and scientists (Warr, 2014). Since the time of inception, many other data sharing initiatives have been set forth. The creation of the Public Library of Science (PLoS) in 2003, has also familiarized researchers with data sharing practices. As an open-access, peer-reviewed journal, PLoS journals have always required all submissions to make the data available to readers (PLoS, 2014). However, both NIH and PLos have had to release new, extended data policies because of lack of participation by researchers or lack of clear direction from NIH and PLoS as to how to share the data (Bloom, 2013). Several issues arise as barriers to data sharing: technological platform, dissemination, accessibility, and storage. However, as technology grows and solves the problems of where, how, and how much data to archive, an over arching obstacle emerges. That obstacle is a cultural issue. There are different cultures based on discipline or scientific area, as the data lifecycle in disciplines vary (Warr, 2014). There is also a problem among scientists concerning ownership, as scientists' strongly believe that the data they have collected belongs to them (Piwowar &Chapman 2010). Another concern expressed by researchers is that they spend time making their data available, yet there is no way to be certain that they will receive any kind of credit for sharing (Warr, 2014).

These cultural issues are currently the largest impedance on sharing data. With these concerns and lack of practice of data sharing policy that PLoS, NIH, and NSF are continuously updating their data sharing policies. For example, PLoS has clarified the requirement of "data available upon request," by providing a platform and eliminating the gatekeeper (Bloom et al, 2014). According to PLoS:

"As of March 1, 2014, authors submitting their research manuscript to PLoS journals will find a field in the online submission form where they will be asked to provide the data availability information, which will then be available to editors and reviewers during the review process and, in the event of publication, will be published with the article." (Bloom et al, 2014).

Now that some journals and government agencies are clarifying their data policies, there should be less grey area as to where, how, and when researchers should share their data. Journals like PLoS are also blazing the trail for other publications to implement similar data sharing policies (Sturges et al, 2014). However, in a 2014 research study on journal data sharing policies, the researchers found that more than half of the journals studied had no data sharing policy (Sturges et al, 2014), suggesting that not only does the culture of researchers need to adapt, but also the culture of publications. A group of researchers at Nottingham University, UK, created a project called JoRD. It was established to "conduct a feasibility study into the scope and shape of a

sustainable service that will collate and summarise journal data policies," (Sturges et al, 2014). The project ran from July to December 2012 and aimed to: "Identify and consult with a wide range of stakeholders, publishers and others, and to develop a detailed set of stakeholder requirements and service specifications (with regard to journal data research)."

This project was an attempt to establish a business framework to provide clear guidance and information related to journal's data sharing policies. It was discovered that a lack of consistency of policy existed, not just amongst different journals, but within the same journal (Sturges et al, 2014). If the scientific journal is going to remain the centerpiece of the communication process of science, journals are going to have to experience a cultural shift, as well (McCain, 1995). Journals are going to have to define exactly what their data sharing policies are in order to get researchers to comply.

**New Researchers vs. Established Researchers**

After conducting the post hoc analyses, it was suggested that newer researchers have a higher endorsement of the norm of reciprocity than their more established counterparts. This finding has particular interest because of the new standards and procedures for sharing data (Holdren, 2013). Perhaps the newer researchers are expressing a stronger endorsement of the norm of reciprocity because they are use to collaborating on research, coming directly out

of academia? The new researchers are from an era where sharing data is a social norm; therefore tit-for-tat would be status quo. However, further post hoc analysis shows that although newer Ph.D.'s report a greater endorsement of the norm of reciprocity, there was no indication in this study's sample that the same group legitimized stronger data sharing practices. What are the expectations of both new and established researchers in regard to data sharing and or the norm of reciprocity?

Another limitation of the research is the dropping of several questions from both measures. It was evident when calculating the reliabilities of the measures that the end items on each scale were problematic. After a few of the items were dropped, the reliabilities of both measures increased. When items are dropped from a measure, some questions will go unanswered. One particular research question of interest that will go unanswered in this study concerns the theory of the norm of reciprocity and states: "If the norm of reciprocity is universal, what will the sharing practices look like in underrepresented groups?" The research question is an important one. Although Gouldner theorized universality of the norm, he did so over 50 years ago. Perhaps our values and acceptance of reciprocity has changed over the last 5 decades?

# LIST OF REFERENCES

Aronson, W. A. (2007). *Social Psychology 6th Edition*. New Jersey: Pearson Education, Inc.

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., et al. (2004). Promoting access to public research data for scientific, economic, and social development. *Data Science Journal, 3:* 135-153.

Axelrod, R. *The Evolution of Cooperation*. Revised edition. New York: Basic Books, 2006.

Baron, R.M. & Kenny, D.A. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology, 51(6),* 1173-1182.

Bello, F. (1960). How to Cope with Information*. Fortune, 162*.

Birnholtz, J. & Bietz, M. (2003). Data at work: supporting sharing in science and engineering. In *Proceedings of the 2003 international ACM SIGGROUP conference on supporting group work*. ACM, New York, NY, USA, 339-348. DOI=10.1145/958160.958215 http://doi.acm.org/10.1145/958160.958215

Bjaalie, J. (2007). The Neuroinformatics Portal of the International Neuroinformatics Coordinating Facility. *Journal of Neuroscience 27*(14): 3613-3615. doi: 10.1523/JNEUROSCI.0558-07.2007.

Bjork, Bo-Christer (2014). Open Access Subject Repositories: An Overview. Journal of the Association for Information Science and Technology (JAIST), 65(4): 698-706.

Bloom, Theo. (2013). Data Access for the Open Access Literature: PLoS's Data Policy, PLoS online. http://www.plos.org/data-access-for-the-open-access-literature-ploss-data-policy/.

Bloom, T., Ganley, E. and Winker, M. (2014). Data Access for the Open Access Literature: PLoS's Data Policy. *PLoS Medicine, 11(2):* e1001607.

Boster, F.J., Baker, C.R., Kingsley, C., Kotowski, M.R., Strom, R.E., & Westerman, D. (2005 November). The Norm of Reciprocity: Simply a favor for a favor or also an eye for an eye? Paper presented at the annual meeting of the National Communication Association, Boston, MA.

Bourne PE, Barbour V (2011) Ten Simple Rules for Building and Maintaining a Scientific Reputation. PLoS Comput Biol 7(6): e1002108. doi:10.1371/journal.pcbi.100210

Campbell, E.G., Clarridge, B. R., Gokhale, M. L., Birenbaum, S., Hilgartner,Holtzman, N. A. & Blumenthal, D. (2002). Data withholding in academic genetics: Evidence from a national survey. *JAMA* 287(4): 473-480.

Carnegie Foundation for the Advancement of Teaching (2014). Carnegie Classifications, Research Universities, Very High Research Activity. http://classifications.carnegiefoundation.org/descriptions/basic.php

Ceci, S. (1988). Scientists' Attitudes toward Data Sharing. *Science, Technology, & Human Values*, 13 (1-2): 45-52.

Chiu, C., Hsu, M., & Wang, E. (2006). Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories. *Decision Support Systems* 42, 1872-1888.

Cialdini, R., Reno, R., & Kallgren, C. (1990). A Focus Theory of Normative Conduct: Recycling the Concept of Norms to reduce Littering in Public Places. *Journal of Personality and Social Psychology,* 58 (6), 1015-1026.

Cialdini, R., Vincent, J., Lewis, S., Catalan, J., Wheeler, D., & Darby, B. (1975). Reciprocal Concessions Procedure for Inducing Compliance: The Door-in-the-Face Technique*. Journal of Personality and Social Psychology, 31*(2): 206-215.

Comte, A. (1868). *The positive philosophy of Auguste Comte*. W. Gowans.

Corbyn, Z. (2011). Researchers Failing to Make Raw Data Public*. Nature: International Weekly Journal of Science,* published online 14 September 2011. doi: 10.1038/news.2011.536

de Solla Price, D. J., de Solla Price, D. J., de Solla Price, D. J., & de Solla Price, D. J. (1986). *Little science, big science... and beyond* (pp. 46-156). New York: Columbia University Press.

Deutsch, M., & Gerard, H.B. (1955). A Study of Normative and Information Social Influences upon Individual Judgment. *Journal of Abnormal and Social Psychology*, 51, 629-636.

Dillman, D. A. (2000). Mail and Internet surveys: The tailored design method (2nd ed.). New York, NY:John Wiley

Douglass, K., Allard, S., Tenopir, C., Wu, L. & Frame, M. (2014). Managing Scientific Data as Public Assets: Data Sharing Practices and Policies among Full-Time government Employees. *Journal of the Association for Information Science and Technology, 65(2)*: 521-262.

Eisenberger, R., Lynch, P., Aselage, J., & Rohdieck, S. (2004). Who Take the Most Revenge? Individual differences in Negative Reciprocity Norm Endorsement. Personality and social Psychology Bulletin, 30 (6): 787-799. DOI: 10.1177/0146167204264047

Foster, N. F. & Gibbons, S. (2005). Understanding Faculty to Improve Content Recruitment for Institutional Repositories. *D-Lib Magazine, 11(1)*, January 2005.

Freedman, J. & Fraser, S. (1966). Compliance without Pressure: The Foot-in-the-Door Technique. *Journal of Personality and Social Psychology, 4*(2), 195-202.

French, J.R.P. (1956). A Formal Theory of Social Power. *Psychological Review, 63*(3), 181.

Galison, P. L., & Hevly, B. W. (Eds.). (1992). *Big Science: The Growth of Large-scale Research; [papers from a Workshop Held at Stanford in 1988]*. Stanford University Press.

Gouldner, A. (1960). The Norm of Reciprocity: A Preliminary Statement. *American Sociological Review, 25(*2), 161-178.

Goodman, A., Pepe, A., Blocker, A. W., Borgman, C. L., Cranmer, K., Crosas, M., Di Stefano, R., Gil, Y., Groth, P., Hedstrom, M., Hogg, D. W., Kashyap, V., Mahabal, A., Siemiginowska, A., & Slavkovic, A. Ten Simple Rules for the Care and Feeding of Scientific Data. PLoS Computational Biology, 10(4): e1003542.

Green, A. G. & Gutmann, M. P. (2007). Building Partnerships among Social Science Researchers, Institution-based Repositories, and Domain Specific Data Archives. OCLC Systems and Services. *International Digital Library Perspectives, 23(1),* 35-53.

Harnad, S. (2006). Generic Rationale for University Open Access Self-Archiving Policy. Available: http://eprints.ecs.soton.ac.uk/12078/01/genericSApolicy-linked.html

Hauesser, C. (2009). A Comparison of Information-Sharing Between Scientists in Academia and the Industry. Proceedings from *Druid Summer Conference, Copenhagen Business School.* Frederiksberg, Denmark, June 17-19, 2009. Retrieved from: http://www2.druid.dk/conferences/viewpaper.php?id=5378&cf=32

Hearings (1956) before a Subcommittee of the Committee on Government Operations, United States Senate, House of Representatives, 84th Congress, Second Session, March 7, 8, and 9, 1956, Availability of Information from Federal Executive Agencies, Part 4-Panel Discussion on Scientific and Technical Information, Friday, March 9, 1956, pp. 876–879.

Hewlett, R.G. & Anderson, O.E. (1962). *The New World, 1939-1946. A History of the United States Atomic Energy Commission.* University Park, Pennsylvania: Pennsylvania State University Press. ISBN 0-520-07186-7.

Hilbert, M. (2011). Mapping the Dimensions and Characteristics of the World's Technological Communication Capacity during the Period of Digitization (1986-2007/2010). ITU Working Paper Presented at the 9th World Telecommunication/ICT Indicators Meeting, Mauritius, 7-9 December 2011. http://www.itu.int/ITU-D/ict/wtim11/documents/inf/015INF-E.pdf

Hilbert, M. & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science, 332(6025)*, 60-65. DOI: 10.1126/science.1200970

Holdren, John P. (2013). Increasing Access to the Results of Federally funded Scientific Research. Memorandum for the Heads of Executive Departments and Agencies. Executive Office of the President. Office of Science and Technology Policy. http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

Iamnitchi, A., Ripeanu, M., & Foster, I. (2002). Locating Data in (Small-World?) Peer-to-Peer Scientific Collaborations. *1st International Workshop on Peer-to-Peer Systems*, Cambridge, MA, March 2002. http://www.cs.rice.edu/Conferences/IPTPS02/172.pdf

Joint Information Systems Committee (2010). How JISC is helping researchers. Available: http://www.jisc.ac.uk/whatwedo/campaigns/res3/jischelp.aspx. Accessed October 2013.

Kaye, J., Heeney, C., Hawkins, N., de Vries, J., & Boddington, P. (2009). Data Sharing in Genomics: Reshaping Scientific Practice. *Nature Reviews Genetics* 10, 331-335. doi:10.1038/nrg2573

Kling, R. & McKim, G. (2000). Not Just a Matter of Time: Field Differences and the Shaping of Electronic Media in Supporting Scientific Communication. *Journal of the American Society for Information Science and Technology, 51:* 1306-1320.

Koslow, S. (2002). Sharing Primary Data: A Threat or Asset to Discovery? *Nature Reviews Neuroscience* 3, 311-313 (April 2002) | doi:10.1038/nrn787.

Kraut, R., Egido, C., & Galegher, J. (1988). Patterns of Contact and Communication in Scientific Research Collaboration. In *Proceedings of the 1988 ACM conference on Computer-supported cooperative work* (CSCW '88). ACM, New York, NY, USA, 1-12. DOI=10.1145/62266.62267 http://doi.acm.org/10.1145/62266.62267

Lederman, Doug (2012). Campus Employment Stagnates. *Inside Higher Ed*, 26 September 2012.

Lippitt, R., Polansky, N., & Rosen, S. (1952). The dynamics of power; a field study of social influence in groups of children. *Human Relations; Human Relations*.

Lynch, C. A. Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. *Libraries and the Academy, 3(2):* 327-336.

Lyman, Peter and Hal R. Varian, "How Much Information", 2003. Retrieved fromhttp://www.sims.berkeley.edu/how-much-info-2003 on August 19, 2013.

Markoff, J. (2013, August 12). How to Share Scientific Data. The New York Times, D3, retrieved from: http://www.nytimes.com/2013/08/13/science/how-to-share-scientific-data.html?pagewanted=all.

Marshall, Gordon & Scott, John, (Eds.). (2009). *The Oxford Dictionary of Sociology*. Oxford: Oxford University Press.

Marwell, G. & Schmitt, D. (1967). Dimensions of Compliance-Gaining Behavior: An Empirical Analysis. *Sociometery,* 30, 350-364.

McCain, K. W. (1995). Mandating Sharing: Journal Policies in the Natural Sciences. *Science Communication, 16:* 403-431.

McMillan,J. H. (2000). Educational research: Fundamentals for the consumer (3rd ed.). New York, NY: Addison Wesley Longman.

Mennes, M., Biswal, B. B., Castellanos, F. X., Milham, M. P. (2013). Making Data Sharing Work: *NeuroImage,* 82: 683-691.

Miller, D.T. & Prentice, D.A. (1996). The Construction of Social Norms and Standards. In F.T. Higgins & A.W. Kruglanski (Eds.), *Social Psychology: Handbook of Basic Principles* (pg. 799-829). New York: Guilford.

Mui, L., Mohtashemi, M. & Halberstadt, A. (2002). A Computational Model of Trust and Reputation. In Proceeding of the 35th Hawaii International Conference on System Sciences. HICSS, Minoa, HI, USA, Vol 7. ISBN:0-7695-1435-9

Murstein, B. I., & Azar, J. A. (1986). The relationship of exchange-orientation to friendship intensity, roommate compatibility, anxiety, and friendship. *Small Group Behavior, 17*, 3-17.

National Institute for Health. Final NIH Statement on Sharing Research Data. [Press Release]. Retrieved from: http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html

National Science Foundation. Grant General Conditions (GC-1). April 1, 2001. Retrieved from: http://www.nsf.gov/pubs/2001/gc101/gc101rev1.pdf

Nelson, B (2009). Data Sharing: Empty Archives. *Nature, 461*, 160-163. doi:10.1038/461160a

Oettinger, A.G. (1965). An Essay in Information Retrieval or The Birth of a Myth. *Information and Control 8(1),* 64-79. http://dx.doi.org/10.1016/S0019-9958(65)90283-4

Panagopoulos, C. (2010). Affect, Social Pressure and Prosocial Motivation: Field Experimental Evidence of the Mobilizing Effects of Pride, Shame, and Publicizing Voting Behavior. *Political Behavior*, 32, 269-386.

Pinfield, S. (2009). Journals and Repositories: An Evolving Relationship? *Learned Publishing, 22(3)*, 165-175, doi:10.1087/2009302.

Piwowar, H. A. and Chapman, W. W. (2010). Public Sharing of Research Datasets: A Pilot Study of Associations. *Journal of Informatics, 4*: 148-156.

PLoS.org. (2014). Site Information.

Postle, B., Shapiro, L., & Biesanz, J. (2002). On Having One's Data Shared. *Journal of Cognitive Neuroscience, 14*(6), 828-840.

Rashid,A., Ling, K., Tassone, R., Resnick, P., Kraut,R., & Riedl, J. (2006). Motivating participation by displaying the value of contribution. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (CHI '06). ACM, New York, NY, USA, 955-958. DOI=10.1145/1124772.1124915 http://doi.acm.org/10.1145/1124772.1124915

Regan, D.T. (1971). Effects of a favor and liking on compliance. *Journal of Experimental Social Psychology, 7*, 627-639.

Registry of Open Access Repositories, ROAR. Web site for open repositories throughout the world. Retrieved: August 19, 2013. roar.eprints.org

Savage, C. J. and Vickers, A. J. (2009). Empirical Study of Data Sharing by Authors Publishing in PLoS Journals. PLoS ONE, 4: e7078. Doi: 10.1371/journal.pone.0007078.

Schofield, P., Bublea, T., Weaver, T., Portilla, L., Brown, S.D., Hancock, J., Einhorn, D., Tocchini-Valentini, G., Hrabe de Angelis, M, Rosenthal, N., & CASIMIR Rome Meeting Participants. (2009). Post-publication Sharing of Data and Tools. *Nature* **461**, 171-173 (10 September 2009) | doi:10.1038/461171a

Stolte, E., von Praun, C., Alonso, G., & Gross, T. (2003). Scientific Data Repositories: Designing for a Moving Target. SIGMOD, June 9-12, 2003. San Diego, CA. Retrieved from: http://www.research.ibm.com/people/p/praun/sigmod03.pdf

Sturges, P., Bamkin, M., Anders, J., Hussain, A. (2014). Access to Research Data: Addressing the Problem through Journal Data Sharing Policies. Proceeding of the IATUL Conferences. Paper 3. http://docs.lib.purdue.edu/iatu./2014/openaccess/3

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A., Wu, L., et al. (2011). Data Sharing by Scientists: Practices and Perceptions. PLos ONE 6(6): e21101.

Väänänen, A., Buunk, B.P., Kivimäki, M., Pentti, J., & Vahtera, J. (2005). When It is Better to Give than Receive: Long-Term Health Effects of Perceived Reciprocity in Support Exchange. *Journal of Personality and Social Psychology, 89(2),* 176-193.

Van Horn, J. D. & Gazzaniga, M. S. (2002). Opinion: Databasing fMRI Studies Towards a 'Discovery Science' of Brain Function. *National Review of Neuroscience, 3:* 314-318.

Van Horn, J. D. & Gazzaniga, M. S. (2013). Why Share Data? Lessons Learned from the fMRIDC. *NeuroImage, 82:* 677-682.

Van Horn, J. D. & Ishai, A. (2007). Mapping the Human Brain: New Insights from FMRI Data Sharing. *Neuroinformatics, 5(2):* 146-153.

Van Horn, J. E., Schaufeli, W. B., & Taris, T. W. (2001). Lack of reciprocity among Dutch teachers: Validation of reciprocity indices and their relation to stress and well-being. *Work & Stress, 3,* 191–213.

Van House, N., Butler, M., & Schiff, L. (1998). Cooperative knowledge work and practices of trust: sharing environmental planning data sets. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work* (CSCW '98). ACM, New York, NY, USA, 335-343. DOI=10.1145/289444.289508 http://doi.acm.org/10.1145/289444.289508

Vines, T. H., Andres, R. L., Bock, D. G., Franklin, M. T., Gilbert, K. J., Kane, N. C., Moore, J., Moyers, B. T., Renault, S., Rennison, D. J., Veen, T., and Yeaman, S.

(2013). Mandated Data Archiving Greatly Improves Access to Research Data. *The FASEB Journal, Life Sciences Forum, 27*: 1304-1307.

Walters, C. D. (2014). The NSF/NIH Effect: Surveying the Effect of Data Management Requirements on Faculty, Sponsored Programs, and Institutional Repositories. *Library Faculty & Staff Publications*, paper 158. http://digitalcommons.usu.edu/lib_pubs/158

Warr, W. A. (2014). Data Sharing Matters. *Journal of Computer Aided Molecular Design, 28:* 1-4.

Whatley, M., Webser, J., Smith, R., & Rhodes, A. (1999). The Effect of a Favor on Public and Private Compliance: How Internalized is the Norm of Reciprocity? *Basic and Applied Social Psychology, 21*(3), 251-259. **DOI:** 10.1207/S15324834BASP2103_8

Xia, Jingfeng. (2008). A Comparison of Subject and Institutional Repositories in Self-archiving Practices. *The Journal of Academic Librarianship, 34(6)*: 489-495.

Zimmerman, A. (2008). New Knowledge from Old Data: The Role of Standards in the Sharing and Reuse of Ecological Data. Science, Technology, & Human Values, 33(5), 631-652. **doi: 10.1177/0162243907306704**

Zimmerman, A. (2003). Data Sharing and Secondary use of Scientific Data: Experiences of Ecologists. (Doctoral Dissertation). Retrieved from Google Scholar, March 15, 2011. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.101.9434&rep=rep1&type=pdf

# APPENDICES

# Appendix A

Measures

Adjusted Eisenberger Scale of Norm of Reciprocity Scale

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | | | | | | | | Strongly Agree |

| | | |
|---|---|---|
| 1. | If someone important to Pat does something negative to Pat, Pat should do something even more negative to them. | 1 2 3 4 5 6 7 |
| 2. | If someone treats Pat badly, Pat should treat them even worse. | 1 2 3 4 5 6 7 |
| 3. | If someone treats Pat badly, Pat should treat that person badly in return. | 1 2 3 4 5 6 7 |
| 4. | Pat should not give help to those who treat others badly. | 1 2 3 4 5 6 7 |
| 5. | If someone dislikes Pat, Pat should dislike them. | 1 2 3 4 5 6 7 |
| 6. | If someone wants to be Pat's enemy, Pat should treat them like an enemy. | 1 2 3 4 5 6 7 |
| 7. | If someone says something nasty to Pat, Pat should say something nasty back. | 1 2 3 4 5 6 7 |
| 8. | If someone distrusts Pat, Pat should distrust them. | 1 2 3 4 5 6 7 |
| 9. | When someone treats Pat badly, Pat should still act nicely to them. | 1 2 3 4 5 6 7 |
| 10. | Pat should feel uncomfortable when someone does him a favor that he knows he won't be able to return. | 1 2 3 4 5 6 7 |
| 11. | If someone does something for Pat, Pat should feel required to do something for them. | 1 2 3 4 5 6 7 |
| 12. | If someone gives Pat a gift, Pat should feel obligated to get them a gift. | 1 2 3 4 5 6 7 |
| 13. | Pat should not forget if someone owes Pat a favor. | 1 2 3 4 5 6 7 |

| | |
|---|---|
| 14. Pat should always repay someone who has done Pat a favor. | 1 2 3 4 5 6 7 |
| 15. If Sam sends Pat a birthday card, Pat should send Sam a birthday card. | 1 2 3 4 5 6 7 |
| 16. If someone does Pat a favor, Pat should repay them in some way. | 1 2 3 4 5 6 7 |
| 17. If someone says something pleasant to Pat, Pat should say something pleasant back. | 1 2 3 4 5 6 7 |
| 18. If someone treats Pat well, Pat should treat that person well in return. | 1 2 3 4 5 6 7 |
| 19. Pat should not forget if Pat owes someone a favor. | 1 2 3 4 5 6 7 |
| 20. If someone is nasty to Pat, Pat should not react. | 1 2 3 4 5 6 7 |

# Appendix B

Data Sharing Scale

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Strongly Disagree | | | | | | Strongly Agree |

| | |
|---|---|
| 1. If Pat is required by a grant funding institute to share data, then Pat should share data. | 1 2 3 4 5 6 7 |
| 2. If funding for Pat's research is contingent on Pat sharing data, then Pat should share data. | 1 2 3 4 5 6 7 |
| 3. Pat should share data in order to promote the scientific method. | 1 2 3 4 5 6 7 |
| 4. Pat should not share data unless Pat is rewarded. | 1 2 3 4 5 6 7 |
| 5. Pat should not share data with researchers unless Pat has a relationship with those researchers. | 1 2 3 4 5 6 7 |
| 6. Pat should share data with colleagues that Pat has collaborated with in the past. | 1 2 3 4 5 6 7 |
| 7. Pat should share data with any researcher that requests Pat's data. | 1 2 3 4 5 6 7 |
| 8. Pat should not share data with anyone requesting Pat's data. | 1 2 3 4 5 6 7 |
| 9. Pat should share data with researchers who are renown in Pat's area of research. | 1 2 3 4 5 6 7 |
| 10. Pat should share data with researchers that have positive reputations. | 1 2 3 4 5 6 7 |
| 11. Pat should not share data with researchers whose reputations are uncertain to Pat. | 1 2 3 4 5 6 7 |
| 12. Pat should share data with researchers who have shared data with Pat. | 1 2 3 4 5 6 7 |
| 13. When Chris shares data with Pat, Pat should | 1 2 3 4 5 6 7 |

share data with Chris.

14. When Chris shares data with Pat, Pat should share data with Jamie. 1 2 3 4 5 6 7

15. If Jamie does not share data with Pat, Pat should not share data with Jamie. 1 2 3 4 5 6 7

16. If Jamie does not share data with Pat, Pat should not share data with Jamie or Chris. 1 2 3 4 5 6 7

17. If Pat gives Chris data, Chris should give Pat data. 1 2 3 4 5 6 7

18. If Pat gives Chris data, Pat should also give Jamie data. 1 2 3 4 5 6 7

19. If Pat gives data to Jamie, Pat should get credit for sharing data. 1 2 3 4 5 6 7

20. If Jamie uses Pat's data, Pat should be acknowledged in Jamie's research. 1 2 3 4 5 6 7

21. If Jamie gives Pat's data to Chris, Chris should acknowledge Jamie and Pat. 1 2 3 4 5 6 7

22. If Jamie gives Pat's data to Chris, Chris should acknowledge Jamie. 1 2 3 4 5 6 7

23. If Jamie gives Pat's data to Chris, Chris should acknowledge Pat. 1 2 3 4 5 6 7

24. If Chris uses Pat's data, Chris does not need to acknowledge Pat 1 2 3 4 5 6 7

# Appendix C

1. What is your age? _____

2. What is your area of research? _____

3. How long have you been conducting research at a university as a faculty member? *(A faculty member is defined as a university employee that conducts research, instructs courses, or both). _____

4. Are you currently receiving funding from an organization that calls for a data management plan for the data collected? _____

5. How many publications have you contributed to since you began conducting research at a university as a faculty member? (A faculty member is defined as a university employee that conducts research, instructs courses, or both) _____(this can be an estimate).

# Appendix D

November 10, 2013
You are invited to participate in an academic research survey especially for scientific researchers. If you are not an active researcher in academe, please feel free to discard this email.
I am a doctoral student at The University of Tennessee. This survey is for my dissertation research on testing the theory of the norm of reciprocity on scientific data sharing. By participating in the survey, you will be contributing valuable insights about your attitudes and behavior as a researcher, which will significantly enhance my understanding of your most important work. The ultimate goal of this research is to gain insight into why scientific data sharing is more of an exception than a rule.
Completing the questionnaire is simple and will only take about 15 minutes of your time. You will receive an email on November 17, 2014 with the survey link.
Your responses will be kept strictly confidential. They will be combined with responses from many other people, solely for the purpose of descriptive statistical analysis.
Additional information about the study is available on the Informed Consent statement, which will is at the bottom of this email and will accompany the survey link. Please complete your survey upon receipt. If you have any questions, please contact me at the email address or phone number listed below. I appreciate your time, effort, and support.
Sincerely,

Crystal Pleake Sherline
University of Tennessee Knoxville
College of Communication and Information
School of Information Science
451 Communication Bldg.
1345 Circle Park Circle
Knoxville, TN 37996-0341
Csherli1@utk.edu
Phone: 865.773.4234

# Appendix E

## INFORMED CONSENT STATEMENT

Project Name: "Testing the Theory of the Norm of Reciprocity on Scientific Data Sharing: The Carrot or the Stick Approach?"

INTRODUCTION

You are invited to participate in a research study about scientific data sharing. The study seeks to gain some understanding of data sharing practices in the academic researcher community. The findings from this research could help improve data management protocol, data repositories, and data sharing practices.

INFORMATION ABOUT THE PARTICIPANTS' INVOLEMENT IN THE STUDY

Your participation in the study involves completing the attached survey questionnaire and submitting it electronically.

Please complete the survey only once. It is estimated that the survey will take no more than 15 minutes to complete.

RISKS

Because participation is limited to completing a survey, there are no foreseeable risks to the participants from their involvement in the study.

BENEFITS

It is anticipated that this research will benefit the participants by extending the body of knowledge about data sharing practices in the participants' discipline.

CONFIDENTIALITY

The information in the study records will be kept confidential. Data will be stored securely and will be made available only to the researcher conducting the study and members of her doctoral committee, unless participants specifically give permission in writing to do otherwise. Data from the survey will only be reported in aggregate terms; no reference will be made in oral or written reports that could link participants to the study.

COMPENSATION

You will not be financially compensated for participating in this study.

EMERGENCY MEDICAL TREATMENT

The University of Tennessee does not "automatically" reimburse participants for medical claims or other compensation. The risk of participating in this study is minimal, so no need for emergency medical treatment is anticipated. If physical injury is suffered during the course of research, or for more information, please notify the investigator:

Crystal Pleake Sherline, College of Communication and Information, at csherli1@utk.edu or (865)773.4234.

CONTACT INFORMATON

If you have questions at any time about the study or procedures, you may contact the researcher, Crystal Pleake Sherline, at The University of Tennessee, College of Communication and Information, 451 Communication Bldg., 1345 Circle Park Drive, Knoxville, TN 37996 and/or (865)773.4234 or email at: csherli1@utk.edu. If you have questions about your rights as a participant, contact the Office of Research Compliance Officer at (865)974.3446.

PARTICIPATION

Your participation in this study is voluntary; you may decline to participate without penalty. If you decide to participate, you may withdraw from the study at any time without penalty and without loss of benefits to which you are otherwise entitled. If you withdraw from the study before data collection is completed your data will be returned to your or destroyed.

# VITA

Crystal Pleake Sherline was born in Port Charlotte, Florida, to the parents of Roger and Sandra Pleake. She is the youngest child and only daughter. She attended Orange River Elementary School, Lee Middle School, and Fort Myers High School. After graduation, she attended the University of Florida, the University of Maryland, Florida State University, and finally, the University of Tennessee. She holds a BS in Biology, a MA in English, a MLIS, and a Ph.D. in Communication and Information. In between her education, she moved to Europe, traveled, had two children, and returned to the US. She currently lectures at the University of Tennessee and works full-time as a technical information analyst for the Homeland Defense and Security Information Analysis Center (HDIAC).